Sven Eberhardt, Jonah Cader & Thomas Serre

How Deep is the Feature Analysis Underlying Rapid Visual Categorization?

We report preliminary rapid categorization results with 20 participants using Amazon Mechanical Turk (psiTurk framework and HTML5 video elements for reliable timing in stimulus presentation across different browsers). 300 images were sampled from ImageNet [3] using the animal category as target and plant and scene categories as distractors. Presentation times were brief (50 ms) and maximum allowed answer time varied between 500 ms and 1500 ms by block (6 blocks with 50 images each).

Feature responses were extracted from different processing stages of the VGG16 (caffe implementation [4] using pre-trained weights). For the convolutional layers, a random subset of 4,096 features were extracted. Model decisions were based on the output of a linear SVM trained on 100,000 samples (C regularization parameter optimized by cross-validation). We computed rank-order correlations between classifier confidence outputs and human accuracy.

Although model accuracy on the animal detection task increases with increasingly deep layers, per-image correlation with human decisions decreases at higher stages (Fig 2). We hypothesize that state-of-the-art hierarchical networks have become “too deep” to correctly model rapid processing in the human visual cortex. We seek to further develop the method of massive online experiments as well as exhaustive search in deep model space towards a data-driven approach for determining receptive field properties underlying rapid human visual processing.

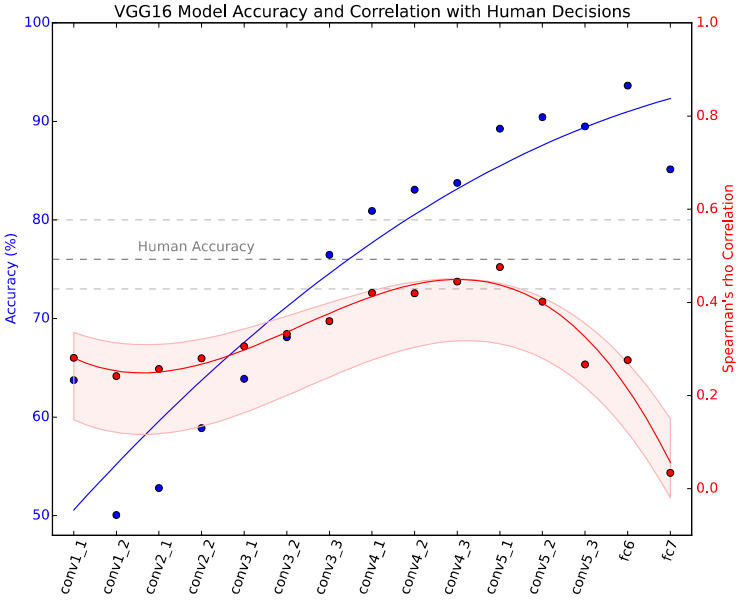

Fig 1. Model accuracy (blue) and per-image rank-order correlation between model and human decisions (red) for different VGG16 layers (horizontal axis). Curves are 2nd and 3rd degree polynomial fits. 95% confidence interval was computed with bootstrapping (1,000 iterations on the subject pool).

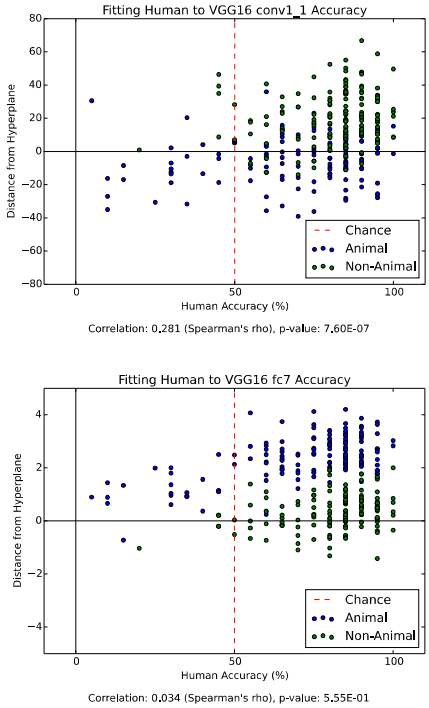

Fig 2. Human versus model decisions for an early (top: conv1_1) and late (bottom: fc7) stage.

Funding: Supported by NSF early career award (IIS-1252951) and DARPA young faculty award (N66001-14-1-4037).

References:

[1] Serre et al. “A feedforward architecture accounts for rapid categorization.” PNAS, 104(15), pp. 6424-6429, 2007

[2] Simonyan et al. "Very deep convolutional networks for large-scale image recognition." *arXiv:1409.1556*, 2014

[3] http://image-net.org

[4] Jia et al. "Caffe: Convolutional architecture for fast feature embedding." *Proc. of the ACM*, 2014